AD-A250 345

REPORT DOCUMENTATION PAGE

Form Approved OPM No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE: 1984

3. REPORT TYPE AND DATES COVERED: Unknown

4. TITLE AND SUBTITLE: Bayesian and Non-Bayesian Evidential Updating

5. FUNDING NUMBERS: DAAB10-86-C-0567

6. AUTHOR(S): Henry E. Kyburg, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES):
University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER: TR-139

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES):
U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER: 92-TRF-0034

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT: Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Four main results are arrived at in this paper. (1) Closed convex sets of classical probability functions provide a representation of belief that includes the representations provided by Shafer probability mass functions as a special case. (2) The impact of "uncertain evidence" can be (formally) represented by Dempster conditioning, in Shafer's framework. (3) The impact of "uncertain evidence" can be (formally) represented in the framework of convex sets of classical probabilities by classical conditionalization. (4) The probability intervals that result from Dempster/Shafer updating on uncertain evidence are included in (and may be properly included in) the intervals that result for Bayesian updating on uncertain evidence.

14. SUBJECT TERMS: Artificial Intelligence, Data Fusion, Bayesian and Non-Bayesion Updating, Evidential Updating, Probability

15. NUMBER OF PAGES: 35

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Bayesian and Non-Bayesian Evidential Updating

Henry E. Kyburg, Jr.

University of Rochester
Rochester, NY 14627

TR 139
July 1984

Abstract

Four main results are arrived at in this paper. (1) Closed convex sets of classical probability functions provide a representation of belief that includes the representations provided by Shafer probability mass functions as a special case. (2) The impact of "uncertain evidence" can be (formally) represented by Dempster conditioning, in Shafer's framework. (3) The impact of "uncertain evidence" can be (formally) represented in the framework of convex sets of classical probabilities by classical conditionalization. (4) The probability intervals that result from Dempster/Shafer updating on uncertain evidence are included in (and may be properly included in) the intervals that result from Bayesian updating on uncertain evidence.

Research for this paper was supported in part by the U.S. Army Electronics Research and Development Command, and was stimulated in large part by conversations with Jerry Feldman and Ron Loui of the Department of Computer Science at the University of Rochester.

BAYESIAN AND NON-BAYESIAN EVIDENTIAL MEASURES*

1. Recent work in both vision systems (Garvey, Wesley) and in knowledge representation (Lowrance, Barnett, Quinlan, Dillard) has employed an alternative, often referred to as Dempster/Shafer updating, to classical Bayesian updating of uncertain knowledge. Various other investigators have gone beyond classical Bayesian conditionalization (MYCIN, EMYCIN, DENDRAL, ...) but in a less systematic manner. It is appropriate to examine the formal relations between various Bayesian and non-Bayesian approaches to what has come to be called evidence theory, in order to explore the question of whether the new techniques are really more powerful than the old, and the question of whether, if they are, this increment of power is bought at too high a price.

2. Classical probability theory supposes (1) that we commence with known statistical distributions, (2) that these distributions are such as to give rise to real-valued probabilities, and (3) that these probabilities can be modified by using Bayes' theorem to conditionalize on evidence that is taken to be certain. There are thus three ways to modify the classical theory.

We may dispense with the supposition that we are dealing with known statistical distributions. The best known advocate of this gambit was L. J. Savage, who argued that probabilities represent personal, subjective, opinions, and not objective distributions of quantities in the world. This approach has given rise to Bayesian statistics, based on the fact that the opinions of most people are such that, faced with frequency data, they will converge reasonably rapidly. Furthermore, in practice, it is common to recognize that some opinions are better than others, and to use as prior distributions in statistical inference distributions representing the opinions of knowledgeable experts. This approach has been incorporated in some expert systems, for example, PROSPECTOR. It has both virtues and limitations. A purely pragmatic virtue is that it allows us to get on with our business even when we don't have the knowledge of prior distributions we would like to have. It has the practical virtue that the considered opinions of genuinely knowledgeable experts are formed in response to, and reflect with some degree of accuracy, relative frequencies in nature. But it has two drawbacks: it does not incorporate any indication of whether the opinion is a wild guess, or a considered judgement based on long experience; and it calls for expert opinions even in the face of total, acknowledged ignorance.

This suggests the second departure from the classical picture; abandoning the assumption that our probabilities are point-valued. This has recently been hailed as a novel departure (Lowrance, 1982, p. 21; Garvey, et. al., 1981, p. 319; Dillard, 1982, p. 1; Lowrance and Garvey, 1982, p.7; Wesley and Hanson, 1982, p. 16; Quinlan, 1982, p. 9). The idea of representing probabilities by intervals is not new (cf. Kyburg, Good, Levi, Smith), and the notion of probabilities that constitute a field richer than that of the real numbers goes back even further (Keynes, 1921, offers a formal philosophical treatment of such entities; B.O. Koopman, 1941, 1942, offers a mathematical characterization). Even the standard subjectivistic or personalist view of probability can be construed in this way; while each person has a set of real valued probabilities defined over a given field, a group of people will reflect a <u>set</u> of probability functions defined over the field. We may quite resonably focus our attention on the supremum and infimum of these functions evaluated at a member of the field.[1]

In general the representation in terms of intervals seems superior to the representation in terms of point values. Even in the <u>ideal</u> case, in which all of our measures are based on statistical inference from suitably massive quantities of data, it is most natural to construe these measures as being constrained by intervals. In confidence interval estimation, for example, what we get from our statistics is a high confidence that a given parameter is contained in a certain interval. This translates neatly and conveniently into an interval constraint. The results of statistical inference should reflect interterminacy or vagueness. What we can properly claim to know is not that a parameter has a certain value, but that it lies within certain limits. This limitation of human knowledge should surely be mirrored in computer based systems.

The third departure from the classical scheme is to consider alternatives to Bayes' theorem as a way of updating probabilities in the light of new evidence. This departure is recent, and was first stated in Dempster, 1967. Dempster's novel rule of combination, subsequently adopted by Shafer (1976), is often referred to as a "generalization" of Bayesian inference (Lowrance and Garvey, 1982, p. 9: "Dempster's rule can be viewed as a direct generalization of Bayes' rule ..."; Dillard, 1982, p.1; Garvey, et. al., p. 319; Lowrance 1982, p. 21). This suggests, on the one hand, that Bayes' rule can be regarded as a special or limiting case of Dempster's rule, which is true, and on the other hand that Dempster's rule can be applied where Bayes' rule cannot, which is false. Dempster himself recognizes (1967, 1968) that his rule results from the imposition of additional constraints on the Bayesian analysis (see note 4).

One very serious problem with the usual Bayesian approach to evidential updating is the quantity of information that must be embodied in the probability function covering the field of propositions with which we are concerned. This may be _empirical_ information (if the underlying probabilities are thought of as being based on statistical knowledge), _psychological_ information (if a personalistic interpretation of probability is adopted), or _logical_ information (if we interpret probability as degree of confirmation, _a la_ Carnap (1951)). Suppose we consider a field of propositions based on the logically independent propositions $p_1 \ldots p_n$; the set of what Carnap called "state descriptions" induced by this basis consists of 2^n atoms, each of which is the conjunction of the n (negated or unnegated) p_i. It is obvious that for reasonably large n this assignment of probabilities presents great difficulties. But once we have those 2^n numbers, we're done - we can calculate all conditional probabilities as well as the probability of any proposition in the field based on $p_1 \ldots p_n$.

Is there a saving in effort if we go to a Dempster/Shafer System? Using the handy representation in Shafer (1976), we take θ, the universal set, to be the set of all 2^n possibilities represented by the state descriptions, and assign a mass to each <u>subset</u> of θ. This requires 2 exp $2^{\underline{n}}$ assignments! As far as the number of parameters to be taken account of is concerned, we are exponentially worse off. But if we construe probabilities as intervals, or represent them by convex sets of simple probability functions, we are just as badly off. (For an example relating mass assignments to interval assignments, see table I in the appendix. For the general equivalence, see theorem 1 below.) Dillard (p. 4) refers to "computational limitations" and Lowrance and Garvey (1982) mention that with large θ, maintaining the model is "computationally infeasible."

In either case, we need to find some systematic and computationally feasible procedure for obtaining the masses or probabilities we need. Bayesian and non-Bayesian approaches are in essentially the same difficult situation in this respect, although there are often plausible ways of systematizing the parameter assignments on either view.

3. Whether the representation of our initial knowledge state is given by an assignment of masses to subsets of θ or by a set of classical probability distributions over the atoms of θ, it is important that these masses or probabilities by justifiable. As already suggested, the most straight-forward way of obtaining them is through statistical

inference, which (when possible) yields interval valued estimates of relative frequencies. But there may also be other ways to obtain masses or intervals of probability. If so, then the deep and difficult problem arises of how to combine both statistical and non-statistical sources of information.[2]

It has been suggested that Dempster/Shafer updating relieves us of the necessity of making assumptions about the joint probabilities of the objects we are concerned about. Thus, Quinlan claims that INFERNO "makes no assumptions whatever about the joint probability distributions of pieces of knowledge ..." (Quinlan 1982). Other writers have made similar claims -- e.g., Wesley and Hanson, 1982, p. 15. (To make independence assumptions is exactly to make assumptions about joint probability distributions.)

It is clear that the assignment of masses to subsets of θ involves just as much in the way of "assumptions" as the assignment of a priori probabilities to the corresponding propositions. In view of the reducibility of the Dempster/Shafer formalism to the formalism provided by convex sets of classical probability functions (to be shown below), moreover, we may recapture the assumptions about joint probability distributions from the convex Bayesian representation.

4. One important novelty of the Dempster/Schafer system is its ability to handle uncertain evidence. But even this is not in itself anti-Bayesian. There are also Bayesian methods for handling uncertain evidence. One of these, used in PROSPECTOR and mentioned by Lowrance (1982, p. 17) is known in the philosophical world as Jeffrey's rule. (It is presented and discussed in Jeffrey, 1965) It follows from Bayes' theorem that

$$\underline{P}(\underline{A}) = \underline{P}(\underline{A}/\underline{B})\underline{P}(\underline{B}) + \underline{P}(\underline{A}/-\underline{B})\underline{P}(-\underline{B}).$$

If you adopt a new (coherent) probability function $\underline{P}'$, there are essentially no constraints on $\underline{P}'(\underline{A})$. But one can adopt the principle that if a shift in probability _originates_ in the assignment of a new probability to $\underline{B}$, that should _not_ affect the _conditional_ probability of $\underline{A}$ _given_ $\underline{B}$: $\underline{P}(\underline{A}/\underline{B}) = \underline{P}'(\underline{A}/\underline{B})$. We have learned something new about $\underline{B}$, but we haven't learned anything new about the bearing of the truth of $\underline{B}$ on the truth of $\underline{A}$.

Given this principle, the response of a shift in the probability of $\underline{B}$ from $\underline{P}(\underline{B})$ to $\underline{P}'(\underline{B})$, resulting from new evidence, should propagate itself according to:

$$\underline{P}'(\underline{A}) = \underline{P}(\underline{A}/\underline{B})\underline{P}'(\underline{B}) + \underline{P}(\underline{A}/-\underline{B})\underline{P}'(-\underline{B})$$

When new evidence leads us to shift our credence in $\underline{B}$ from $\underline{P}(\underline{B})$ to $\underline{P}'(\underline{B})$, a corresponding shift in probability is induced for every other proposition in the field: the new probability of a proposition $\underline{A}$ is the weighted average of the probability of $\underline{A}$, given $\underline{B}$, and the probability of $\underline{A}$ given not-$\underline{B}$, weighted by the new probabilities of $\underline{B}$ and not-$\underline{B}$.

Lowrance (1982) worries about the problem of iterating this move. Having made it, should we then update the probability of $\underline{B}$ in the light of the new probability $\underline{P}'(\underline{A})$? Wesley and Hanson (1982, p. 15) worry about a potential "violation of Bayes' law." But what is offered is not a relaxation method: it is a method of evaluating the impact of evidence which warrants a shift in the support for B. It makes no sense to consider updating $P'(B)$ in the light of the new value of $P(A)$; $P'(B)$ is the source of the updating. No contradiction lurks here.

Other Bayesian updating procedures are possible (cf. Hartry Field, 1978), but it is hard to think of one so simple and so natural. This is particularly true in the epistemological framework considered by Shafer; the weights of the subsets of θ assigned masses reflect our a priori intuitions; there is no way in which the values of these masses, given our observations, can be changed without changing the model entirely. What impact given evidence has should not also change according to the evidence we happen to have.

5. . In order to investigate more closely the relations between the Bayesian and Dempster/Shafer stragegies for updating, it will be helpful to have several formal results. In the present section we establish the partial equivalence between the assignment of masses to subsets of θ (the space of possibilities) and the assignment of a convex set of simple classical probability functions defined over the atoms of θ. The equivalence is only partial, since some plausible situations do not have a representation in terms of mass functions.[3] (Throughout "$\subset$" is to be understood as proper or improper inclusion.)

Theorem 1:

Let m be a probability mass function defined over a frame of discernment θ. Let $\underline{Bel}(\underline{X})$ be the corresponding belief function -- $\underline{Bel}(\underline{X}) = \sum_{\underline{A} \subset \underline{X}} \underline{m}(\underline{A})$. Then there is a closed, convex set of classical probability functions $\underline{S}_p$ defined over the atoms of θ such that for every subset $\underline{X}$ of $\underline{\theta}$, $\underline{Bel}(\underline{X}) = \inf_{\underline{P} \in \underline{S}_p} \underline{P}(\underline{X})$

Proof: Let $\underline{S}_p$ be the set of classical probability functions $\underline{P}$ defined on the atoms of θ such that for every $\underline{X} \subset \theta$, $\underline{Bel}(\underline{X}) \leq \underline{P}(\underline{X}) \leq 1-\underline{Bel}(\underline{X})$. $\underline{S}_p$ is closed, since $\underline{P}(\underline{X}) = \underline{Bel}(\underline{X})$, $\underline{P}(\overline{\underline{X}}) = 1-\underline{Bel}(\underline{X})$ is a classical probability function. $\underline{S}_p$ is convex, since for $0 < \underline{a} < 1$, $\underline{aP}_1(\underline{X}) + (1-\underline{a})\underline{P}_2(\underline{X})$ lies between $\underline{Bel}(\underline{X})$ and $1-\underline{Bel}(\underline{X})$ whenever $\underline{P}_1(\underline{X})$ and $\underline{P}_2(\underline{X})$ do. Since there is a $\underline{P} \in \underline{S}_p$ such that $\underline{P}(X) = \underline{Bel}(\underline{X})$, $\underline{Bel}(\underline{X}) \geq \inf_{\underline{P} \in \underline{S}_p} \underline{P}(\underline{X})$. And $\inf_{\underline{P} \in \underline{S}_p} \underline{P}(\underline{X}) \geq \underline{Bel}(\underline{X})$ since this inequality holds for every $\underline{P} \in \underline{S}_p$.

Theorem 2

If $\underline{S}_{\underline{P}}$ is a closed convex set of classical probability functions defined over the atoms of θ, and for every $\underline{A}$, $\underline{B} \subset \theta$, inf $\underline{P}(\underline{A} \cup \underline{B}) \geq$ inf $\underline{P}(\underline{A})$ + inf $\underline{P}(\underline{B})$ - inf $\underline{P}(\underline{A} \cap \underline{B})$, then there is a mass function $\underline{m}$ defined over the subsets of θ such that for every $\underline{X}$ in θ, the corresponding Bel function satisfies

$$\underline{\text{Bel}}\ (\underline{X}) = \inf_{\underline{P} \in \underline{S}_{\underline{P}}} \underline{P}(\underline{X}) \ .$$

proof: Since $\underline{S}_{\underline{P}}$ is closed and convex, for every $\underline{X} \subset \theta$ there is a $\underline{P} \in \underline{S}_{\underline{P}}$ such that $\underline{P}(\underline{X}) = \inf_{\underline{P} \in \underline{S}_{\underline{P}}} \underline{P}(\underline{X})$. For every $\underline{X} \subset \theta$, define $\underline{P}_{\star}(\underline{X})$ to be $\inf_{\underline{P} \in \underline{S}_{\underline{P}}} (\underline{X})$.

By Shafer's Theorem 2.1, if θ is a frame of discernment then a function Bel $2^{\theta} \to [0,1]$ is a belief function if and only if

(1) $\underline{\text{Bel}}\ (\emptyset) = 0$ $\qquad$ $\underline{P}_{\star}(\emptyset) = 0$

(2) $\underline{\text{Bel}}\ (\theta) = 1$ $\qquad$ $\underline{P}_{\star}(\theta) = 1$

(3) For every positive integer $\underline{n}$ and every collection $\underline{A}_1, \ldots, \underline{A}_n$ of subsets of θ,

$$\underline{\text{Bel}}\ (\underline{A}_1 \cup \ldots \cup \underline{A}_{\underline{n}}) \geq \sum_{\underline{I} \subseteq \{1,\ldots,\underline{n}\}} (-1)^{|\underline{I}|+1}\ \underline{\text{Bel}}\Big(\bigcap_{\underline{i} \in \underline{I}} \underline{A}_{\underline{i}}\Big)$$

Since Shafer's theorem 2.2 gives an algorithm to recapture the mass function from the belief function, we need merely extablish (3) for our function $\underline{P}_{\star}$:

$$(3') \quad \underline{P}_*(\underline{A}_1 \cup \cdots \underline{A}_{\underline{n}}) \geq \sum_{\underline{I} \subset \{1,\ldots,\underline{n}\}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}})$$

Suppose (3') fails. Then there is a collection $\underline{A}_1,\ldots,\underline{A}_{\underline{n}}$, of smallest cardinality $\underline{n}$, for which (3') is false. I.e.,

$$\underline{P}_*(\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}}) < \sum_{\underline{I} \subset \{1,\ldots,\underline{n}\}} (-1)^{|\underline{I}|+1} P_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}})$$

But $\underline{P}_*(\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}}) \geq \underline{P}_*(\underline{A}_{\underline{n}}) + \underline{P}_*(\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}-1}) - \underline{P}_*(\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}-1}) \cap \underline{A}_{\underline{n}})$,

by the hypothesis of the theorem.

$$\underline{P}_*((\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}-1}) \cap \underline{A}_{\underline{n}}) = \underline{P}_*((\underline{A}_1 \cap \underline{A}_{\underline{n}}) \cup (\underline{A}_2 \cap \underline{A}_{\underline{n}}) \cup \cdots \cup (\underline{A}_{\underline{n}-1} \cap \underline{A}_{\underline{n}}))$$

By hypothesis, (3') holds for collections of cardinality of $(\underline{n}-1)$.

$$(4) \quad \text{Thus } \underline{P}_*((\underline{A}_1 \cap \underline{A}_{\underline{n}}) \cup (\underline{A}_2 \cap \underline{A}_{\underline{n}}) \cup (\underline{A}_{\underline{n}-1} \cap \underline{A}_{\underline{n}})) \geq \sum_{\underline{I} \subset \{1,\ldots,\underline{n}-1\}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}} \cap \underline{A}_{\underline{n}})$$

$$(5) \quad \text{and } \underline{P}_*(\underline{A}_1 \cup \cdots \cup \underline{A}_{\underline{n}-1}) \geq \sum_{\underline{I} \subset \{1,\ldots,\underline{n}-1\}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}})$$

Compute $\displaystyle\sum_{\underline{I} \subset \{1,\ldots,\underline{n}\}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}})$:

We evaluate the sum by cases: $|\underline{I}| = 1$, $|\underline{I}| > 1$ and $\underline{n} \notin \underline{I}$, and $|\underline{I}| > 1$ and $\underline{n} \in \underline{I}$.

$$|\underline{I}| = 1 : \qquad \underline{P}_*(\underline{A}_{\underline{n}}) + \sum_{\substack{\underline{I} \subset \{1,\ldots,\underline{n}-1\} \\ |\underline{I}|=1}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}})$$

$$= \sum_{\substack{\underline{I} \subset \{1,\ldots,\underline{n}\} \\ |\underline{I}|=1}} (-1)^{|\underline{I}|+1} \underline{P}_*(\bigcap_{\underline{i}\in\underline{I}} \underline{A}_{\underline{i}}) = \underline{P}_*(\underline{A}_{\underline{n}}) + \sum_{\underline{i}<\underline{n}} \underline{P}_*(\underline{A}_{\underline{i}})$$

$|\underline{I}| > 1,\ \underline{I}\subset\{1,\ldots,\underline{n}-1\}:\ \sum_{\underline{I}\subset\{1,\ldots,\underline{n}-1\}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}})$

$|\underline{I}| > 1,\ \underline{I}'\ \{1,\ldots,\underline{n}-1\}:\qquad \underline{I} = \underline{I}'\cup\{\underline{n}\},\ \underline{I}'\subset\{1,\ldots,\underline{n}-1\}$

$$\sum_{\underline{I}'\subset\{1,\ldots,\underline{n}-1\}} (-1)^{|\underline{I}'|+2}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}^1} \underline{A}_{\underline{i}}\cap\underline{A}_{\underline{n}})$$

$$= \sum_{\substack{\underline{I}\subset\{1,\ldots,\underline{n}\} \\ \underline{n}\varepsilon\underline{I},\ |\underline{I}|>1}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}})$$

Combining the three terms, we have,

$$P_*(\underline{A}\cup\ldots\cup\underline{A}_{\underline{n}}) \geq \sum_{\substack{\underline{I}\subset\{1,\ldots,\underline{n}\} \\ |I| > 1}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}}) + \sum_{\substack{\underline{I}\subset\{1,\ldots,\underline{n}\} \\ \underline{n}\notin\underline{I} \\ |\underline{I}|>1}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}})$$

$$+ \sum_{\substack{\underline{I}\subset\{1,\ldots,\underline{n}\} \\ \underline{n}\varepsilon\underline{I} \\ |\underline{I}|>1}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}})$$

$$= \sum_{\underline{I}\subset\{1,\ldots,\underline{n}\}} (-1)^{|\underline{I}|+1}\ \underline{P}_*(\bigcap_{\underline{i}\varepsilon\underline{I}} \underline{A}_{\underline{i}})$$

These two theorems show that the representation of uncertain knowledge provided by Shafer's probability mass functions is exactly equivalent to a representation provided by a convex set of classical probability functions, and that the representation of uncertain knowledge by a convex set of classical probability functions is exactly equivalent to a representation provided by a probability mass function so long as the convex set of probability functions satisfies the general relation $P_*(A\cup B) \geq P_*(A) + P_*(B) - P_*(A\cap B)$.

6. The main theorem of this section gives the relation between convex Bayesian updating and Dempster/Shafer updating. To establish the theorem requires two reductions. These are given by two lemmas. The first provides an algorithm for computing the result of Dempster/Shafer updating in response to uncertain evidence; the second does the same thing for Bayesian updating.

Lemma 1: Let θ be a frame of discernment. Let our initial belief function by $\underline{Bel}_1$. We obtain new evidence whose impact on the frame of discernment θ can be represented by a simple support function (Shafer 1976, p. 7) $\underline{Bel}_{\underline{C}}$ whose single focus is $\underline{C} \varepsilon 2^{\theta}$. $\underline{Bel}_{\underline{C}}$ attributes mass $\underline{s}$ to $\underline{C}$ and mass $(1-\underline{s})$ to θ.

Let the foci of $\underline{Bel}_1$ - the subsets $\underline{A}$ of θ receiving mass $m_1(\underline{A})>0$ - be $\underline{A}_1, \underline{A}_2, \ldots, \underline{A}_{\underline{n}}$. We can construct a new frame of discernment θ' and a new belief function $\underline{Bel}_1'$, such that

(a) For every $\underline{X} \subseteq \theta$, $\underline{Bel}_1'(\underline{X}) = \underline{Bel}_1(\underline{X})$

(b) For every $\underline{X} \subseteq \theta$, $(\underline{Bel}_1 \oplus \underline{Bel}_{\underline{C}})(\underline{X}) = \underline{Bel}_1'(\underline{X}|\underline{E})$, where $\underline{E} \varepsilon 2^{\theta'}$, and the evidence partially supporting $\underline{C}$ provides total support for $\underline{E}$. "$\oplus$" represents the application of Dempster's rule of combination to $\underline{Bel}_1$ and $\underline{Bel}_{\underline{C}}$; $\underline{Bel}_1'(\underline{X}|\underline{E})$ represents Dempster's rule of conditioning on $\underline{E}$ - the analog of Bayesian conditionalization (Shafer 1976 p. 67).

Proof: Let $\underline{e}$ be new to θ, and for every $\underline{p}\varepsilon\theta$ generate two new "possibilities" $\underline{pe}$ and $\underline{p\bar{e}}$. Let $\theta' = \{\underline{p}': \exists \underline{p}\varepsilon\theta(\underline{p}'=\underline{pe} \vee \underline{p}'=\underline{p\bar{e}})\}$. Let $\underline{E} = \{\underline{p}': \exists\underline{p}\varepsilon\theta(\underline{p}'=\underline{pe})\}$. Since the evidence that supports $\underline{C}$ is to render $\underline{E}$ certain, we have $\underline{C}'\subset\underline{E}$ i.e. $\underline{C}' = \{\underline{p}': \exists\underline{p}\varepsilon\underline{C}(\underline{p}'=\underline{pe})\}$.

We define $\underline{Bel}_1'$ on the basis of $\underline{m}_1'$ as follows:

$\underline{Bel}_1'$ has $\underline{n}$ foci of the form $\underline{A}_{\underline{i}}$, each with mass $(1-\underline{s})\underline{m}_1(\underline{A}_{\underline{i}})$, where $\underline{m}_1$ is the mass function associated with $\underline{Bel}_1$.

For every $\underline{i}$ such that $\underline{A}_{\underline{i}}\cap\underline{C}' = \emptyset$, $\underline{A}_{\underline{i}}\cap\underline{\bar{E}}$ is to be a focus with mass $\underline{s}\cdot\underline{m}_1(\underline{A}_{\underline{i}})$. For convenience we take the first $\underline{p}$ of the $\underline{A}_{\underline{i}}$ to be those for which $\underline{A}_{\underline{i}}\cap\underline{C}' = \emptyset$. Note that $\underline{p}$ may be 0, but cannot be $\underline{n}$, else $\underline{Bel}_1 \oplus \underline{Bel}_{\underline{C}}$ would be undefined.

The remaining $\underline{i}$ give rise to the remaining foci. These are of the form $(\underline{A}_{\underline{i}}\cap\underline{C}')\cup(\underline{A}_{\underline{i}}\cap\underline{\bar{E}})$, and receive the remaining mass. Since $(\underline{A}_{\underline{i}}\cap\underline{C}')\cup(\underline{A}_{\underline{i}}\cap\underline{\bar{E}})=(\underline{A}_{\underline{j}}\cap\underline{C}')\cup(\underline{A}_{\underline{j}}\cap\underline{\bar{E}})$ is a possibility for $\underline{i}\neq\underline{j}$, we write

$$\underline{m}_1'((\underline{A}_{\underline{i}}\cap\underline{C}')\cup(\underline{A}_{\underline{i}}\cap\underline{\bar{E}})) = \sum_{\{\underline{j}:\ (\underline{A}_{\underline{j}}\cap\underline{C}')\cup(\underline{A}_{\underline{j}}\cap\underline{\bar{E}}) = (\underline{A}_{\underline{i}}\cap\underline{C}')\cup(\underline{A}_{\underline{i}}\cap\underline{\bar{E}})\}} \underline{m}_1(\underline{A}_{\underline{j}})\cdot\underline{s}$$

Note that $\sum\underline{m}_1'((\underline{A}_{\underline{i}}\cap\underline{C}')\cup(\underline{A}_{\underline{i}}\cap\underline{\bar{E}})) = \sum_{\underline{i}=\underline{p}+1}^{n} \underline{m}_1(\underline{A}_{\underline{i}})\cdot\underline{s}$, since

these sets have positive mass only if $\underline{A}_{\underline{i}}\cap\underline{C}'\neq\emptyset$.

We first show that $\underline{Bel}_1'$ is a belief function. Obviously its mass function $\underline{m}'$ is non-negative for every $\underline{A} \subseteq \theta'$, so we need only show that

$\sum_{\underline{A} \subseteq \theta'} \underline{m}'(\underline{A}) = 1$. Summing over the three kinds of foci, we have:

$$\sum_{\underline{A} \subseteq \theta'} \underline{m}'(\underline{A}) = \sum_{\underline{i}=1}^{\underline{n}} (1-\underline{s})\underline{m}_1(\underline{A}_{\underline{i}}) + \sum_{\underline{i}=1}^{\underline{p}} \underline{s} \cdot \underline{m}_1(\underline{A}_{\underline{i}}) + \sum_{\underline{i}=\underline{p}+1}^{\underline{n}} \underline{s} \cdot \underline{m}_1(\underline{A}_{\underline{i}}) = 1.$$

We next show that $\underline{Bel}_1'$ is equivalent to $\underline{Bel}_1$ - i.e. that for any $\underline{X} \subseteq \theta$, $\underline{Bel}_1'(\underline{X}) = \underline{Bel}_1(\underline{X})$.

$$\underline{Bel}_1'(\underline{X}) = \sum_{\underline{A} \subseteq \underline{X}} \underline{m}'(\underline{A}) = \sum_{\underline{A}_{\underline{i}} \subseteq \underline{X}} \underline{m}'(\underline{A}_{\underline{i}}) + \sum_{\underline{A}_{\underline{i}} \cap \underline{E} \subseteq \underline{X}} \underline{m}'(\underline{A}_{\underline{i}} \cap \underline{E}) + \sum_{(\underline{A}_{\underline{i}} \cap \underline{C}) \cup (\underline{A}_{\underline{i}} \cap \overline{\underline{E}}) \subseteq \underline{X}} \underline{m}'((\underline{A}_{\underline{i}} \cap \underline{C}) \cup (\underline{A}_{\underline{i}} \cap \overline{\underline{E}}))$$

The first term yields $\sum_{\underline{A}_{\underline{i}} \subseteq \underline{X}} (1-\underline{s})\underline{m}_1(\underline{A}_{\underline{i}}) = (1-\underline{s}) \sum_{\underline{A}_{\underline{i}} \subseteq \underline{X}} \underline{m}_1(\underline{A}_{\underline{i}})$

Since $\underline{X} = (\underline{X} \cap \underline{E}) \cup (\underline{X} \cap \overline{\underline{E}})$, $\underline{A}_{\underline{i}} \cap \overline{\underline{E}} \subseteq \underline{X} \cap \overline{\underline{E}}$ if and only if $\underline{A}_{\underline{i}} \subseteq \underline{X}$, in view of the fact that $\underline{p}\overline{\underline{e}} \in \underline{A}_{\underline{i}} \cap \overline{\underline{E}}$ if and only if $\underline{p}\overline{\underline{e}} \in \underline{A}_{\underline{i}}$, and the same holds for $\underline{X}$. Thus the second term yields $\sum_{\substack{\underline{A}_{\underline{i}} \subseteq \underline{X} \\ 1 \leq \underline{i} \leq \underline{p}}} \underline{s} \cdot \underline{m}_1(\underline{A}_{\underline{i}})$

To evaluate the third term, we claim that $(A_i \cap C) \cup (A_i \cap \overline{E}) \subseteq X$ if and only if $A_i \subseteq X$. If $A_i \subseteq X$, then $A_i \cap C \subseteq X$ and $A_i \cap \overline{E} \subseteq X$ and so $(A_i \cap C) \cup (A_i \cap \overline{E}) \subseteq X$. Suppose $(A_i \cap C) \cup (A_i \cap \overline{E}) \subseteq X$. Then $A_i \cap \overline{E} \subseteq X$, $A_i \cap \overline{E} \subseteq X \cap \overline{E}$, and by the preceding argument $A_i \subseteq X$. Thus the third term yields $\sum_{\substack{A_i \subseteq X \\ p < i \leq n}} s \cdot m_1(A_i)$

Putting the three parts together, we have $Bel_1'(X) = Bel_1(X)$.

We now show that conditioning on E in the frame of discernment θ' is equivalent to combining uncertain evidence C with Bel_1 in the frame of discernment θ according to Dempster's rule of combination:

For every $X \subseteq \theta$, $(Bel_1 \oplus Bel_C)(X) = Bel_1'(X \mid E)$

$$(1) \quad (Bel_1 \oplus Bel_C)(X) = \frac{\sum\limits_{A_i \cap C \neq 0,\ A_i \cap C \subseteq X} m_1(A_i) \cdot s \; + \; \sum\limits_{A_i \subseteq X} m_1(A_i)(1-s)}{1 - \sum\limits_{A_i \cap C = \emptyset} m_1(A_i) \cdot s}$$

(The numerator comprises two sums, since Bel_C has two foci: C and θ with masses s and $(1-s)$ respectively.)

$$(2) \quad Bel_1'(X \mid E) = \frac{\sum\limits_{A \subseteq X \cup \overline{E}} m_1'(A) - \sum\limits_{A \subseteq \overline{E}} m_1'(A)}{1 - \sum\limits_{A \subseteq \overline{E}} m_1'(A)}$$

$\sum_{A \subset \bar{E}} m_1'(A) = \sum_{i=1}^{p} s \cdot m_1(A_i)$, since only the foci of the form $A_i \cap \bar{E}$ are included in $\bar{E}$: $A_i = (A_i \cap E) \cup (A_i \cap \bar{E})$ is not included in $\bar{E}$, and since $C' \subset E$, $(A_i \cap C') \cup (A_i \cap \bar{E})$ is included in $\bar{E}$ only if $A_i \cap C' = \emptyset$, in which case it has no mass.

$$\sum_{A_i \cap C' = \emptyset} m_1(A_i) \cdot s = \sum_{i=1}^{p} s \cdot m_1(A_i).$$

Hence the denominators of (1) and (2) are the same.

It remains to evaluate $\sum_{A \subset X \cap \bar{E}} m_1'(A)$. Consider foci of the form A_i. $A_i \subset X \cup \bar{E}$ if and only if $A_i \subset X$, so these foci yield mass

$$\sum_{A_i \subset X} m'(A_i) = \sum_{A_i \subset X} (1-s) m_1(A_i)$$

corresponding to the right hand term in the numerator of (1).

Consider foci of the form $A_i \cap \bar{E}$. All of these are included in $X \cup \bar{E}$; they yield

$$\sum_{i=1}^{p} m'(A_i) = \sum_{i=1}^{p} s \cdot m_1(A_i) = \sum_{A \subset \bar{E}} m_1'(A).$$

So they drop out of the numerator of (2).

Finally, consider foci of the form $(A_i \cap C') \cup (A_i \cap \overline{E})$. We first show that $(A_i \cap C') \cup (A_i \cap \overline{E}) \subset X \cup \overline{E}$ if and only if $A_i \cap C' \subset X$. Suppose $(A_i \cap C') \cup (A_i \cap \overline{E}) \subset X \cup \overline{E}$. Then $A_i \cap C' \subset (X \cup \overline{E})$. But $C' \subseteq E$, so $A_i \cap C' = A_i \cap C' \cap E \subset X \cup \overline{E}$ only if $A_i \cap C' \subset X$. Suppose $A_i \cap C' \subset X$. Then since $A_i \cap \overline{E} \subset \overline{E} \subset X \cup \overline{E}$, $(A_i \cap C') \cup (A_i \cap \overline{E}) \subset X \cup \overline{E}$.

We compute the mass in the numerator of (2) due to foci of this sort. They have mass only when $A_i \cap C' \neq 0$. And then they have mass

$$\sum_{\{j:\ (A_j \cap C') \cup (A_j \cap \overline{E}) = (A_i \cap C') \cup (A_i \cap \overline{E})\}} s \cdot m_1(A_i) \qquad ;$$

each A_i such that $A_i \cap C' \subset X$ contributes $s \cdot m_1(A_i)$. Their total mass is therefore

$$\sum_{\substack{A_i \cap C' \subset X \\ A_i \cap C' \neq 0}} s \cdot m_1(A_i) \,,$$

corresponding to the first term of the numerator of (1).

We have therefore shown that $(Bel_1 \oplus Bel_C)(X) = Bel_1'(X\ E)$.

Two remarks on this construction are in order. First, we have given no rule for finding the "possibility" E. But in general that should be no problem. Suppose C is the proposition that there is a squirrel on the roof of the barn. The light is bad, so $\underline{Bel}_C$ assigns a mass of only .8 to C, and assigns the remaining mass to θ. We take E in θ' to be the proposition that it seems (.8) to be the case that there is a squirrel on the roof, for which the evidence is conclusive. The index 0.8 indicates the force of the seeming, and is reflected in our assignment of masses in θ'. In many situations it seems quite natural to replace "uncertain evidence" by the "certain" data on which it is based.

Second, however, whether or not we can always do this is unimportant for the comparison of Bayesian and Dempster conditioning. We can regard the introduction of E to be merely a computational device that helps us to compare the distribution of masses in θ according to the function $Bel_1 \oplus Bel_C$ to the corresponding set Bayesian conditional distributions.

We now present an analogous Lemma for Bayesian conditionalization based on Jeffrey's rule for uncertain evidence.[4]

Lemma 2.

Suppose that P_0 is our original assignment of probabilities to the field F of propositions whose basis is $a_1, a_2, \ldots a_n$. As a result of stimulation of our sense organs, or unreliable observation, we shift our probability assigned to $\underline{A}$ from $\underline{P}_0(\underline{A})$ to $\underline{P}_1(\underline{A})$. By Jeffrey's Rule, for $\underline{X}\varepsilon\underline{F}$,

$$\underline{P}_1(\underline{X}) = \underline{P}_0(\underline{X}|\underline{A})\cdot\underline{P}_1(\underline{A}) + \underline{P}_0(\underline{X}|\overline{\underline{A}})\cdot\underline{P}_1(\overline{\underline{A}})$$

Let us add a new atomic proposition $\underline{e}$ to the basis of $\underline{F}$ to obtain the field $\underline{F}'$, and represent it by $\underline{E}$. We impose the constraint $\underline{P}_0'(\underline{A}|\underline{E}) = \underline{P}_1(\underline{A})$; $\underline{P}_0'(\underline{E})$ may have any value that strikes our fancy.

We extend $\underline{P}_0'$ so that for any $\underline{X}\varepsilon\underline{F}$, $\underline{P}_0'(\underline{X}) = \underline{P}_0(\underline{X})$; $\underline{P}_0'$ is fully equivalent to $\underline{P}_0$, so far as $\underline{F}$ is concerned, before we obtain information about $\underline{A}$. Specifically, set

$$\underline{k} = \frac{\underline{P}_1(\underline{A})}{\underline{P}_0(\underline{A})}; \quad \underline{k}' = \frac{\underline{P}_1(\overline{\underline{A}})}{\underline{P}_0(\overline{\underline{A}})} = \frac{1-\underline{P}_1(\underline{A})}{1-\underline{P}_0(\underline{A})} = \frac{1-\underline{k}\ \underline{P}_0(\underline{A})}{1-\underline{P}_0(\underline{A})}$$

For $\underline{X}\varepsilon\underline{F}'$, set

$$\underline{P}_0'(\underline{X}\wedge\underline{E}) = \underline{P}_0'(\underline{E})\cdot[\underline{k}\ \underline{P}_0(\underline{X}\wedge\underline{A}) + \underline{k}'\ \underline{P}_0(\underline{X}\wedge\overline{\underline{A}})]$$

$$\underline{P}_0'(\underline{X}\wedge\overline{\underline{E}}) = \underline{P}_0(\underline{X}) - \underline{P}_0'(\underline{X}\wedge\underline{E})$$

Clearly, for $\underline{X}\varepsilon\underline{F}$,

$$\underline{P}_0'(\underline{X}) = \underline{P}_0'(\underline{X}\wedge\underline{E}) + \underline{P}_0'(\underline{X}\wedge\overline{\underline{E}}) = \underline{P}_0(\underline{X})$$

We now show that for $\underline{X}\varepsilon\underline{F}$, probabilities conditional on $\underline{E}$ are equal to the probabilities given by Jeffrey's rule: $\underline{P}_1(\underline{X}) = \underline{P}_0'(\underline{X}|\underline{E})$.

For $\underline{X}\varepsilon\underline{F}$, $\underline{P}_0'(\underline{X}|\underline{E}) = \dfrac{\underline{P}_0(\underline{X}\wedge\underline{E})}{\underline{P}_0'(\underline{E})}$

$$= \frac{\underline{P}_0'(\underline{E})\cdot[\underline{k}\ \underline{P}_0(\underline{X}\wedge\underline{A}) + \underline{k}'\ \underline{P}_0(\underline{X}\wedge\overline{\underline{A}})]}{\underline{P}_0'(\underline{E})}$$

$$= \frac{\underline{P}_0(\underline{X}\wedge\underline{A})}{\underline{P}_0(\underline{A})}\ \underline{P}_1(\underline{A}) + \frac{\underline{P}_0(\underline{X}\wedge\overline{\underline{A}})}{\underline{P}_0(\overline{\underline{A}})}\ \underline{P}_1(\overline{\underline{A}}) = \underline{P}_1(\underline{X})$$

The same remarks may be made with regard to this construction as were made with regard to the previous one. Although we haven't given a rule for specifying E, it shouldn't be too hard in most circumstances to come up with a plausible E; and in any event we can construe the construction as a computational device to make it easier to compare Dempster conditioning and Bayesian conditionalization.

The following theorem shows that in the case of certain evidence, Dempster/Shafer updating yields narrower probability intervals than does Bayesian updating. The next theorem employes Lemmas 2 and 3 to show that this relation holds in general, and not only when our evidence is certain.

Theorem 3:

Let Θ *be a frame of* discernment, Bel a belief function, and S_P the corresponding set of Bayesian probability functions. Let B be evidence assigned probability 1, or support 1. Then for $A \subseteq \Theta$,

$$\inf_{P \in S_P} P(A|B) \leq Bel(A|B) \leq P^*(A|B) \leq \sup_{P \in S_P} P(A|B)$$

where $P^*(A|B) = 1 - Bel(\bar{A}|B)$ is Shafer's plausibility function.

Proof: (All infima and suprema are taken over $P \in S_P$.)

$$\inf P(A|B) = \frac{\inf(A \cap B)}{\inf(A \cap B) + \sup(\bar{A} \cap B)}$$

$$\sup P(A|B) = \frac{\sup(A \cap B)}{\sup(A \cap B) + \inf(\bar{A} \cap B)}$$

$$Bel(A|B) = \frac{Bel(A \cup \bar{B}) - Bel(\bar{B})}{1 - Bel(\bar{B})}$$

$$P^*(A|B) = \frac{P^*(A \cap B)}{P^*(B)} = \frac{1 - Bel(\bar{A} \cup \bar{B})}{1 - Bel(\bar{B})}$$

By computations from table I of the appendix, we obtain:

$$\underline{\inf}\ \underline{P}(\underline{A}|\underline{B}) = \frac{\underline{X}_1}{(\underline{X}_1+\underline{X}_3)+(\underline{X}_{13}+\underline{X}_{23}+\underline{X}_{34})+(\underline{X}_{123}+\underline{X}_{134}+\underline{X}_{234})+\underline{X}_\theta}$$

$$\underline{Bel}(\underline{A}|\underline{B}) = \frac{\underline{X}_1+[\underline{X}_{12}+\underline{X}_{14}+\underline{X}_{124}]}{(\underline{X}_1+\underline{X}_3)+(\underline{X}_{13}+\underline{X}_{23}+\underline{X}_{34})+(\underline{X}_{123}+\underline{X}_{134}+\underline{X}_{234})+\underline{X}_\theta+[\underline{X}_{12}+\underline{X}_{14}+\underline{X}_{124}]}$$

$$\underline{P}^*(\underline{A}|\underline{B}) = \frac{\underline{X}_1+(\underline{X}_{12}+\underline{X}_{13}+\underline{X}_{14})+(\underline{X}_{123}+\underline{X}_{124}+\underline{X}_{134})+\underline{X}_\theta}{(\underline{X}_1+\underline{X}_3)+(\underline{X}_{12}+\underline{X}_{13}+\underline{X}_{14})+(\underline{X}_{123}+\underline{X}_{124}+\underline{X}_{134})+\underline{X}_\theta+[\underline{X}_{23}+\underline{X}_{34}+\underline{X}_{234}]}$$

From which the inequalities easily follow.

Corollary:

(1) $\underline{\inf}\ \underline{P}(\underline{A}|\underline{B}) = \underline{Bel}(\underline{A}|\underline{B})$ iff $\underline{X}_{12}+\underline{X}_{14}+\underline{X}_{234} = 0$

(2) $\underline{Bel}(\underline{A}|\underline{B}) = \underline{P}^*(\underline{A}|\underline{B})$ iff $\underline{X}_{13}+\underline{X}_{123}+\underline{X}_{134}+\underline{X} = 0$

(3) $\underline{\sup}\ \underline{P}(\underline{A}|\underline{B}) = \underline{P}^*(\underline{A}|\underline{B})$ iff $\underline{X}_{23}+\underline{X}_{34}+\underline{X}_{234} = 0$

Theorem 4: If we apply Dempster's rule of combination to any evidence represented by a separable support function (our initial state need not be so represented) we obtain constraints more severe than those we get from Bayesian conditionalization applied to the same initial state.[5]

Proof: A seperable support function may be represented as the combination of simple support functions. By Lemma 1, the effect of a simple support function can be represented by Dempster conditioning. By Theorem 2, the initial state can be represented by a closed convex set of Bayesian probability functions. By Lemma 2 the effect of uncertain evidence (as reflected by a simple support function) can be represented by Bayesian conditionalization. By Theorem 3 the belief intervals resulting from Bayesian conditionalization will include the belief intervals obtained from Dempster conditioning. Therefore the result of applying Dempster's rule of conditioning will lead to belief intervals more severely constrained than the convex Bayesian intervals corresponding to them.

8. Dempster/Shafer evidential updating, we have seen, leads to more tightly constrained representations of rational belief than does convex Bayesian updating.[6] It might be thought that this is a virtue. But whether or not this is a Good Thing is open to question.

Suppose that $\underline{D} = \underline{D}_1, \ldots, \underline{D}_n$ are alternative decisions open to you, and that you have a utility function defined over the cross product of $\underline{D}$ and the set θ of possible states. You begin with a belief function, and you obtain some evidence. If you combine this evidence with your initial belief function according to convex Bayesian conditionalization, your new beliefs will be characterized by a set of probability functions $\underline{P}_B$. If you perform the combination of evidence according to non-Bayesian procedures, your new beliefs will be characterized by a set of probability functions $\underline{P}_N$ that is (in general) a proper subset of $\underline{P}_B$.[7]

Given any probability function $\underline{P}$ in either $\underline{P}_B$ of $\underline{P}_N$, you can calculate the expected value of each decision: $\underline{E}(\underline{D}_i, \underline{P})$. Let us say that $\underline{D}_i$ is <u>admissible</u> relative to a set of probability functions just in case there is some probability function in the set according to which the expected value of $\underline{D}_i$ is at least as great as the expected value of any other decision.[8] Since $\underline{P}_N$ is included in $\underline{P}_B$, the admissible decisions we obtain if we update in a non-Bayesian way are included among those we obtain if we update in a Bayesian way.

There are three cases to consider. (1) We obtain the same set of admissible decisions by either updating procedure. In this case we have gained nothing. (2) If $\underline{P}_N$ leads to a set of admissible decisions containing more than one member, then so does $\underline{P}_B$, and we must in either case invoke additional constraints in order to generate a unique decision. (3) If $\underline{P}_N$ leads to a unique admissible decision and $\underline{P}_B$ does not, we appear to have accomplished something useful by means of non-Bayesian updating.

But it is open to question whether the added power should be built into the evidential updating rule, or whether it should appear as part of a decision procedure that takes us beyond the evidence. Many people feel that principles of evidence and principles of decision should be kept distinct.

Consider an urn filled with black and white iron balls, some of which are magnetized and some of which are not. It is easy to imagine that by extensive sampling, or by word of the manufacturer, our statistical knowledge about the contents of the urn may be as represented in table II of the appendix, where the set of black balls is represented by $\underline{A}$, and the set of magnetized balls is represented by $\underline{B}$. Given that this is our initial state, we may ask what our attitude should be toward the proposition that a ball selected from the urn is magnetic, given that it is white.

Dempster conditioning yields the degenerate interval [0.8, 0.8]

Bayesian conditionalization yields the interval [0.5, 0.8]

Suppose you are offered a ticket for $ .75 that returns a dollar if the ball is magnetic. On the view identified with Dempster and Shafer, it is not only permissible, but, given the usual utility function, <u>mandatory</u> to buy it. On the convex Bayesian view either accepting or rejecting the offer would be admissible. It is true that, <u>for all you know</u>, the true expectation is positive; but it is also true, <u>for all you know</u>, the true expectation is negative. If every thing you know is true, the expected loss may still be $-.25.

On the other hand, there are cases where Dempster's rule of combination leads to intuitively appealing results, but the convex Bayes approach does not.[9] Suppose you know that 70% of the soft berries in a certain area are good to eat, and that 60% of the red berries are good to eat. What are the chances that a soft red berry is good to eat? The rule yields .42/.54 = .78, which has intuitive appeal. But the set of distributions compatible with the conditions of the problem leaves the probability of a soft red berry being good to eat completely undetermined: it is the entire interval [0,1]! It is possible that 100% of the soft red berries are good, and it is possible that 0% of the soft red berries are good.

It is clear that in applying the rule of combination, we are implicitly constraining the set of (joint) distributions we regard as possible. This is suggested by Shafer's requirement that the items of evidence to be combined be "distinct" or "independent". The most natural sufficient condition that leads to the same result as Dempster's rule of combination is that all the probability functions in our convex set satisfy the three conditions

(i) $P(G) = \frac{1}{2}$

(ii) $P(S/G\&R) = P(S/G)$

(iii) $P(S/\overline{G}\&R) = P(S/\overline{G})$.

Condition (i), of course, is our old friend, the principle of indifference. Conditions (ii) and (iii) might be called inverse conditional independence, and it is not hard to imagine that we have warrent for supposing they are satisfied.

The exact necessary and sufficient conditions for agreement between the two methods are that our set of probability functions satisfy one of the two conditions

(iv) $P(\overline{G}\&R\&S)/P(G\&R\&S) = P(\overline{G}\&R)*P(\overline{G}\&S)/P(G\&R)*P(G\&S)$

or (v) $P(S/\overline{G}\&R)/P(S/\overline{G}) = P(\overline{G})/P(G) * P(S/G*R)/P(S/G)$

If our evidence is statistical in character, it clearly behooves us to unpack the statistical assumptions underlying our employment of non-Bayesian updating procedures. But what if our evidence is <u>not</u> statistical in character?

One plausible response is that Dempster's rule of combination is not designed for all cases in which you have statistical data to serve as input. Sometimes the masses in the belief function are determined by frequencies, and sometimes they are not; only when they are <u>not</u> determined by frequencies should we apply non-Bayesian updating. It is difficult to make a case against this response except by making a case for the claim that all responsible and useful probabilities, even very vague one, are based on statistical knowledge. But then we must also face the problem of how to treat evidence which is mixed -- which contains both statistical components and intuitive components. While it is a theorem that Dempster combination is both commutative and associative, and also a theorem that Bayesian combination is both commutative and associative, it is obviously not the case that a mixture of Dempster and Bayesian methods need be commutative and associative.

It should be strongly emphasized that the present arguments are not intended as arguments in favor of the general applicability of convex Bayesian conditionalization. Rather, what I have shown is (1) that the representation of belief states by distributions of masses over subsets of a set θ of possibilities is a special case of the convex Bayesian representation in terms of simple classical probabilities over the atoms of θ, (2) that the treatments of uncertain evidence in both Bayesian and non-Bayesian updating are reducible to the corresponding treatments of certain evidence, and (3) that non-Bayesian updating yields more determinate belief states as outcomes, but that the benefits afforded by non-Bayesian updating are limited and questionable.

Table 1

Atomic propositions A,B

		Mass	Lower Measure	Upper Measure
①	AB	X_1	X_1	$1-X_2-X_3-X_4-X_{23}-X_{24}-X_{34}-X_{234}$
②	$A\bar{B}$	X_2	X_2	$1-X_1-X_3-X_4-X_{13}-X_{14}-X_{34}-X_{134}$
③	$\bar{A}B$	X_3	X_3	$1-X_1-X_2-X_4-X_{12}-X_{14}-X_{24}-X_{124}$
④	$\overline{AB}$	X_4	X_4	$1-X_1-X_2-X_3-X_{12}-X_{13}-X_{23}-X_{123}$
	① ∪ ②	X_{12}	$X_1+X_2+X_{12}$	$1-X_3-X_4-X_{34}$
	① ∪ ③	X_{13}	$X_1+X_3+X_{13}$	$1-X_2-X_4-X_{24}$
	① ∪ ④	X_{14}	$X_1+X_4+X_{14}$	$1-X_2-X_3-X_{23}$
	② ∪ ③	X_{23}	$X_2+X_3+X_{23}$	$1-X_1-X_4-X_{14}$
	② ∪ ④	X_{24}	$X_2+X_4+X_{24}$	$1-X_3-X_1-X_{13}$
	③ ∪ ④	X_{34}	$X_3+X_4+X_{34}$	$1-X_1-X_2-X_{12}$
	① ∪ ② ∪ ③	X_{123}	$X_1+X_2+X_3+X_{12}+X_{13}+X_{23}+X_{123}$	$1-X_4$
	① ∪ ② ∪ ④	X_{124}	$X_1+X_2+X_4+X_{12}+X_{14}+X_{24}+X_{124}$	$1-X_3$
	① ∪ ③ ∪ ④	X_{134}	$X_1+X_3+X_4+X_{13}+X_{14}+X_{34}+X_{134}$	$1-X_2$
	② ∪ ③ ∪ ④	X_{234}	$X_2+X_3+X_4+X_{23}+X_{24}+X_{34}+X_{234}$	$1-X_1$
	θ	X_θ	1	1

$X_\theta = 1-\Sigma X_i$

Table II

A: white B: magnetic

	Mass		Frequency
$\underline{AB}$	X_1	0.2	[0.2,0.4]
$\underline{A\bar{B}}$	X_2	0.2	[0.2,0.4]
$\underline{\bar{A}B}$	X_3	0.1	[0.1,0.2]
$\underline{\bar{A}\bar{B}}$	X_4	0.2	[0.2,0.5]
	X_{12}	0.1	[0.4,0.7]
	X_{13}	0.0	[0.2,0.5]
	X_{14}	0.1	[0.4,0.7]
	X_{23}	0.0	[0.3,0.5]
	X_{24}	0.1	[0.4,0.7]
	X_{34}	0.0	[0.3,0.5]
	X_{123}	0.0	[0.6,0.8]
	X_{124}	0.0	[0.8,0.9]
	X_{134}	0.0	[0.6,0.8]
	X_{234}	0.0	[0.6,0.8]
	θ	0.0	[1.0,1.0]

1. This approach is similar to that of Smith (1961). It is also similar to the approach of Levi (1974, 1981) Good (1962), and Kyburg (1974), but as Levi points out in (1981) there are important differences. Levi represents a credal state by a closed convex set of conditional probability functions. Since distinct closed convex sets of conditional probability functions give rise to the same closed convex sets of simple probability functions (probabilities conditional on tautological evidence), the two representations are not equivalent. Smith and Kyburg represent a credal state by the convex closure of all probabilities consistent with a set of probability intervals. Shafer, as will be seen, implicitly offers the same characterization. Dempster (1968) offers a more restricted characterization: the convex set representing the credal state is the largest that both satisfies the interval constraints, <u>and</u> can be obtained from a space of "simple joint propositions" in a certain way. Levi has shown (1981, pp. 338-392) that these additional restrictions are incompatible with certain natural forms of direct inference of probabilities from known statistics.

2. In another place I shall argue that we can found all our probabilities on direct or indirect statistical inference, or on set-theoretical truths. No other source is needed.

3. An example suggested in conversation by Teddy Seidenfeld is this: consider a compound experiment consisting of either tossing a fair coin twice, or drawing a coin from a bag containing 40% double headed and 60% double tailed coins. The two parts of the compound are performed in an unknown ratio. Let A be the event that the first toss lands heads and B the event that the second toss lands heads. The representation by a

convex set of probability functions is straight-forward, but

$$P_*(A\ B) = 0.75 < 0.9 = P_*(A) + P_*(B) - P_*(A\ B) = 0.4 + 0.5 - 0.0$$

By theorem 2.1 of Shafer 1976, P_* is therefore not a belief function. It is possible to compute a mass function, but the masses assigned to the union of any three atoms must be negative.

4. This result was stated informally by Levi (1967).

5. Dempster (1967, 1968) was well aware that his rule of combination led to results stronger than those that would be given by a mere generalization of Bayesian inference. His reasons for preferring the rule at which he arrives are essentially philosophical: in a classical Bayesian framework, unless you restrict the family of priors, you don't get useful results starting with 0 information. But in expert systems, we have no desire or need to start with zero information.

6. Quinlan's (1982) subtitle suggests the opposite: "A cautious approach to uncertain inference."

7. It is not clear that Shafer's belief functions were intended to be used in a decision-theoretic context. Even if they were, there would be serious difficulties standing in the way of such employment. (See Levi (1978, 1980, 1983), and Seidenfeld (1978)). For present purposes, these difficulties need not concern us.

8. This corresponds to Levi's notion (1981) of E-admissibility.

9. This elegant and simple example was proposed by Jerry Feldman.

References

Barnett, J.A., (1981), "Computational Methods for a Mathematical Theory of Evidence," Proceeding IJCAI 7, 868-875.

Buchanan, B.G., G.L. Sutherland and E.A. Feigenbaum, (1969), "Heuristic Dendral: A Program for Generating Explanatory Hypotheses in Organic Chemistry," Machine Intelligence, Meltzer and Michie (eds), Edinburgh University Press, Edinburgh.

Carnap, Rudolf, (1950), The Logical Foundations of Probability, University of Chicago Press, Chicago.

Dempster, Arthur P., (1967), "Upper and Lower Probabilities Induced by a Multi-valued Mapping," Annals of Mathematical Statistics 38, 325-339.

Dempster, Arthur P., (1968), "A Generalization of Bayesian Inference," Journal of the Royal Statistical Society, Series B, 30, 205-247.

Dillard, R.A., (1982), "The Dempster-Shafer Theory Applied to Tactical Fusion in an Inference System," Fifth MIT/ONR Workshop.

Duda, R.O., P.E. Hart and N.J. Nilsson, (1976), "Subjective Bayesian Methods for Rule-Based Inference Systems," Proceedings 1976 National Computer Conference 45, 1075-1082.

Field, Hartry, (1978), "A Note on Jeffrey Conditionalization," Philosophy of Science 45, 361-367.

Garvey, R.D., J.D. Lowrance and M.A. Fishler, (1981), "An Inference Technique For Integrating Knowledge from Disparate Sources," Proceedings IJCAI 7, 319-325.

Good, I.J., (1962), "Subjective Probability as a Measure of a Non-Measurable Set," Nagel, Suppes and Tarski (eds), Logic, Methodology and Philosophy of Science.

Jeffrey, Richard, (1965), The Logic of Decision, McGraw-Hill, New York.

Keynes, J.M., (1921), A Treatise on Probability, MacMillan and Co., London.

Koopman, Bernard O., (1940), "The Bases of Probability," Bulletin of the American Mathematical Society 46, 763-774.

Koopman, Bernard O., (1940), The Axioms and Algebra of Intuitive Probability," Annals of Mathematics 41, 269-292.

Koopman, Bernard O., (1941), "Intuitive Probabilities and Sequences," Annals of Mathematics 42, 169-187.

Kyburg, Henry E., Jr., (1961), Probability and the Logic of Rational Belief, Wesleyan University Press, Middletown, Ct.

Levi, Isaac, (1974), "On Inderterminate Probabilities, " Journal of Philosophy 71, 391-418.

Levi, Isaac, (1967), "Probability Kinematics," British Journal for the Philosophy of Science, 18, pp. 197-209.

Levi, Isaac, (1979), "Dissonance and Consistency According to Shackle and Shafer," PSA 1978, Asquith and Hacking (eds), Philosophy of Science Association, 466-477.

Levi, Isaac, (1980), "Potential Surprise: Its Role in Inference and Decision Making," in Cohen and Hesse (eds), Applications of Inductive Logic, Oxford University Press, 1-27.

Levi, Isaac, (1981), The Enterprise of Knowledge, the MIT Press, Cambridge MA.

Levi, Isaac, (1983), "Consonance, Dissonance and Evidentiary Mechanism," in Gardenfors, Hansson, and Sahlin (eds), Evidentiary Value, Library of Theoria #15, 27-43.

Lowrance, John E., (1982), "Dependency-Graph Models of Evidential Support, Technical Report, University of Massachusetts, Amherst, MA.

Lowrance, John D., and Thomas D. Garvey, (1982), "Evidential Reasoning: A Developing Concept," IEEE 1982 Proceedings of the International Conference on Cybernetics and Society, 6-9.

Mackie, J.S., (1969), "The Relevance Criterion of Confirmation," British Journal for the Philosophy of Science 20, 27-40.

Quinlan, J.R., (1982), Inferno: A Cautious Approach to Uncertain Inference, A Rand Note, California.

Savage, L.J., (1954), The Foundations of Statistics, John Wiley and Sons, New York.

Seidenfeld, Teddy, (1979), "Statistical Evidence and Belief Functions," PSA 1978, Asquith and Hacking (eds), Philosophy of Science Association, pp. 451-465.

Shafer, Glenn, (1976), A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ.

Shortliffe, E.H., (1976), Computer-Based Medical Consultations: MYCIN, American Elsevier, New York.

Smith, C.A.B., (1961), "Consistency in Statistical Inference and Decision," Journal of the Royal Statistical Society, Series B 23, 1-37.

Smith, C.A.B., (1965), "Personal Probability and Statistical Analysis," Journal of the Royal Statistical Society, Series A, 128, 469-499.

Wesley, Leonard P., and Allen R. Hanson, (1982), "The Use of an Evidential Based Model for Representing Knowledge and Reasoning about Images in the Vision System," IEEE, 14-25.